Anomaly Detection With High Resolution Hyperspectral Observations

ANOMALY DETECTION WITH HIGH RESOLUTION HYPERSPECTRAL OBSERVATIONS

Cécile Chenot, Mehrdad Yaghoobi, Mike E. Davies

School of Engineering
Institute for Digital Communications
The University of Edinburgh
Kings Buildings, Edinburgh, U.K.

Yoann Altmann

School of Engineering and Physical Sciences
Heriot-Watt University
Edinburgh, U.K.

ABSTRACT

Hyperspectral images enable the detection of targets due to the high spectral sampling. The latest generation of sensors also provides an unprecedented spatial resolution which is further exploited in this article to uncover hard to detect anomalies. In particular, we model and estimate the background building upon robust supervised linear unmixing. We benefit from the high resolution of the data to spatially constrain the background. This provides a novel framework for exploiting both the spectral and the energy variations created by the presence of unknown targets to detect them.

Index Terms— Hyperspectral imaging, anomaly detection, linear mixture model.

1. INTRODUCTION

Some of the latest hyperspectral sensors can now be easily mounted on a UAV or a plane due to their compactness and low weight [2]. Besides, they offer a high spatial resolution of the area which enables the detection of small targets (unexpected materials), for defence or civilian applications.

In many hyperspectral analysis tasks, it is assumed that the observation process is mostly linear [3], and data are well described by the Linear Mixture Model (LMM). In this case, it is assumed that the observations $\mathbf{X} \in \mathbf{R}_+^{m\times t}$, where m designates the number of spectral bands and t the total number of pixels, correspond to a sum of n elementary contributions. Each of them is associated with a unique material i and is assumed to be rank-1: it can be factorized as the product between its spectral signature $\mathbf{A}^i \in \mathbf{R}_+^{m\times 1}$ and its spatial distribution $\mathbf{S}_i \in \mathbf{R}_+^{1\times t}$.
This model can be recast in the following matrix form: $\mathbf{X} = \mathbf{AS}+\mathbf{N}$, where $\mathbf{A} \in \mathbf{R}_+^{m\times n}$ designates the endmembers, $\mathbf{S} \in \mathbf{R}_+^{n\times t}$ the abundances, and $\mathbf{N} \in \mathbf{R}^{m\times t}$ represents the noise term, usually assumed to be Gaussian, accounting for perturbations and model imperfections.
Moreover, it is also generally assumed that the abundances $\mathbf{S}$ belong to the simplex $\mathcal{S}^{n-1}$: $\sum_{i=1}^n \mathbf{S}_i^k = 1, \forall k = 1..t$.

We will assume that some man-made targets are also present in the observed scene. They will be modelled by an extra linear term $\mathbf{O} \in \mathbf{R}^{m\times t}$. Since the targets are rare, we can assume that $\mathbf{O}$ is column sparse. We will used the following model to represent the observations:

$$\mathbf{X} = \mathbf{AS} + \mathbf{O} + \mathbf{N},$$

where $\mathbf{A}$, $\mathbf{S}$ are non-negative and $\mathbf{A}$ represents the known endmembers of the background and $\mathbf{S}$ the associated abundances. The simplex prior on $\mathbf{S}$ will not be necessarily enforced. This model has been also presented in various settings [4], [5] and [6] as some examples.

Contributions In this article, we aim to uncover man-made targets by detecting anomalies in the residual between the background, whose endmembers $\mathbf{A}$ are known, and the observations. For this purpose, we will start by discussing the detection and false alarm properties of some basic background modellings and associated priors. It will appear that none of the standard pixel-wise priors can enable the retrieval of hard to detect targets while being robust to energy and illumination variabilities.
To overcome this issue, we will further introduce some spatial regularizations in the LMM to improve the background modelling. In particular, we will present two approaches: the spatial regularization of abundances $\mathbf{S}$ with the LMM and an expanded model of the abundances, spatially constrained as well, in the spirit of the Extended LMM presented in [7].

2. GEOMETRIC INSIGHTS

The aim of this article is to investigate how to detect the outliers and not to recover the abundances (this task is slightly different). Moreover, we simplify the detection problem by looking for the background modelling leading to the best detection properties without the need to estimate jointly the background and outliers.

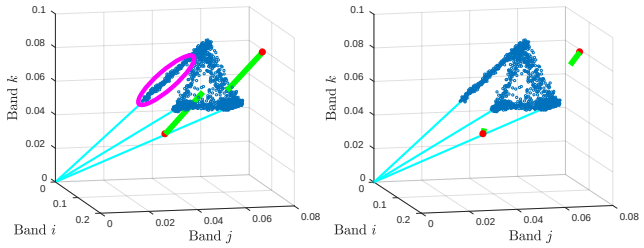
In the following, we will use the simple example presented in Fig.1 to illustrate the problem. The background is composed of 3 materials. We are in the presence of 2 targets. We also model energy and illumination variations for all components (scaling effect for reflectance data [7]).

2.1. Detecting the outliers

In order to model the background, we can exploit one of these geometrical models: the cone model ($\mathbf{S}$ is non-negative) and the simplex model ($\mathbf{S}$ further belongs to $\mathcal{S}^{n-1}$). We discuss briefly the interest of these two options for outliers detection:

- **Simplex model:** This assumption is common in hyperspectral unmixing [3]. Since the energy of each component is bounded, we

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) [Grant number P/K014277/1]; and the MOD University Defence Research Collaboration (UDRC) in Signal Processing. Y.A. is supported by the Royal Academy of Engineering - Research Fellowship scheme (RF201617/16/31). The authors would like to thank A.Halimi for providing his results for [1] and the DSTL for the data.



(a) Simplex (b) Cone

Fig. 1: 3D scatter plots of $\mathbf{X}$ at the bands i, j, and k illustrating the effects of the priors (cone/simplex) on the outlier estimation. The blue points denote the background samples, the red samples the outliers. The blue lines symbolize the edges of the cone generated by $\mathbf{A}$. The green arrows symbolize the estimated outliers (only for the red samples). The samples in the magenta ellipse in (a) would be estimated as outliers.

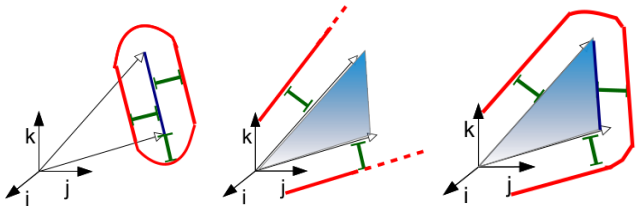
can separate the anomalies and the background using the energy of the components in addition to their spectral difference, see Fig.2-(a) and the difference between Fig.1-(a) and (b). On the other hand, we cannot deal with illumination and energy variations for the same reason: in Fig.1 (a), most of the shadowed pixels in the magenta ellipse would be estimated as being outliers leading to a high false alarm rate. A shadow endmember can be added such as [8] in order to generate a capped cone Fig.2-(c) to cope with shadowing effects but the detection of the anomalies having a smaller energy than the background becomes challenging.
Last, we mention that we cannot enforce the non-negativity of the outliers with the simplex prior on the background: the residual between the observation sample and any of its projection on the simplex generated by $\mathbf{A}$ is non-positive whenever the observation sample (the target material) is less energetic than the background materials.

- **Cone model:** The abundances are only assumed to be non-negative. A cone model prior allows for energy variations. This more flexible model can mostly detect the anomaly based on the spectral variation introduced by the target, Fig.2-(b). In contrast to the simplex model, it struggles to differentiate correlated samples, even if the energy of the man-made target is different from the background: in Fig.1 (b), one can notice that one target is hard to detect since it is very close to the cone generated by $\mathbf{A}$ (anomaly spectrum correlated with $\mathbf{A}$) .

A summary of the key properties of the two models is presented in Table 1.

2.2. Limitations of the standard models

Methods based on the simplex prior have generally a higher true detection rate since they exploit both the energy and spectral diversities between the background and the targets. As a downside, they expose a higher false alarm since every pixel associated with a significant energy variation, such as a shadowed pixel, is estimated as being an outlier. In order to keep a good true detection rate and maintain the false alarm, we can exploit spatial information to state whether a pixel close to the cone generated by $\mathbf{A}$ but not the simplex, belongs or not to the background.



(a) Simplex (b) Cone (c) Capped cone

Fig. 2: Illustration of the different models with 3d scatter plot at the bands i, j and k of $\mathbf{X}$ and $n = 2$. The points in blue represent the expected background. The points on the red surface have the same probability to be an outlier (with an energy-deviation based hypothesis testing).

Model	True Positive	False Positive
Simplex	+	- Too sensitive to energy variations of the background.
Cone	- Cannot detect targets whose spectra are correlated with the background.	+

Table 1: Summary of the properties of the two background models (+ indicates good properties).

On the other hand, the cone model is too flexible and only relies on the spectral difference between the samples and the background material spectra $\mathbf{A}$. As well, we can exploit the neighbouring information to constrain the abundance samples in order to lessen the influence of the targets on the estimated background.

3. SPATIAL INFORMATION

In this section, we will present two methods, based upon the simplex and the cone models, to improve the modelling of the background exploiting the high spatial resolution of the data to spatially constrain the background.

3.1. Cone model and regularization of the abundances

One way to exploit the spatial information is to assume that the background abundances are spatially structured and almost piecewise constant. Indeed, if the spatial resolution of the data is high, then the variations from one pixel to another are quite small and large variations (boundaries between different components) are rare. We can then exploit this morphological diversity between the outliers (rare and sparse) and the abundances (piece-wise constant), to efficiently detect the outliers [9].

This can be achieved by using a Total Variation on $\mathbf{S}$ [10], [11] without the simplex prior, which is too sensitive to energy variations. This can be interpreted as a local simplex prior assumption, enforcing the abundances to be locally constant and allowing for energy variations.

It is worth enforcing that promoting such a spatial structure of the abundances can be quite challenging since if they are over-smoothed, then a large residue will appear and will be estimated as being an outlier. We will limit these artefacts by introducing a weighting scheme in order to regularized mostly the anomalous samples.
This weighting scheme aims to lessen the influence of the samples which are outlying the cone generated by $\mathbf{A}$. The joint weighting and spatial regularization can be interpreted as a relaxed inpainting of the corrupted entries [12], [13]. This is not vital but leads to more robustness because we do not have to over-smooth the components in the absence of outliers.
We estimate the background by solving the following problem:

$$\underset{\mathbf{S}\geq 0}{\operatorname{argmin}} \frac{1}{2} \left\|(\mathbf{X}-\mathbf{AS})\mathbf{W}_O\right\|_2^2 + \sum_{i=1}^{n} \alpha_i \left\|\mathbf{S}_i\Phi\right\|_1 \qquad (1)$$

The weighting matrix $\mathbf{W}_O \in \mathbf{R}^{t\times t}$ is a diagonal matrix, whose elements are given by $\mathbf{W}_{O\,k,k} = \frac{\gamma}{\epsilon+\left\|\mathbf{X}^k-\mathbf{A}\tilde{\mathbf{S}}^k\right\|_2}$, $\forall k = 1..t$ where ϵ and γ are some numerical constants for numerical stability and $\tilde{\mathbf{S}}$ correspond to the non-negative projection of the observations on the cone. Thus, it is based on the outlier estimates with the cone model. Matrix Φ denotes the transformed domain (wavelets, DCT, reformulation of the TV norm etc.) in which the abundances are sparse.
The results would be slightly improved if re-estimating the weights according to the current estimates of $\mathbf{S}$ but at the cost of increasing the computational load. Consequently, we choose to estimate only once the weighting matrix $\mathbf{W}_O$ and to penalize the samples having a large spectral difference with the background.

The algorithm is described as the following: i) estimate $\tilde{\mathbf{S}}$ by assuming that it is non-negative, ii) initialize $\mathbf{W}_O$ with $\tilde{\mathbf{S}}$ and iii) estimate $\mathbf{S}$ with eq.(1) with a Condat-Wu implementation [14].

The spatial resolution may be not high enough and it is then challenging to regularize directly the abundances, which vary too strongly from one pixel to another. In such situations, it would be safer to consider the other method proposed in the following.

3.2. Simplified Extended Linear Mixture Model

We propose another approach for modelling the background, by simplifying the Extended Linear Mixture Model (ELMM) presented in [7]. We assume that the abundances lie in a simplex which can shift in the cone defined by $\mathbf{A}$ so as to compensate energy variations. The background is modelled with $\mathbf{AS}\Lambda$, where $\mathbf{S}$ lie in $\mathcal{S}^{n-1}$ and $\Lambda \in \mathbf{R}^{t\times t}$ is a diagonal matrix. The diagonal elements of $\mathbf{\Lambda}$ can account for the variations of energy by deviating from 1. These shifts will be spatially regularized in our formulation. This model, as well as the ELMM, simply expands the abundance term in the LMM and is equivalent to the cone model without the spatial regularization.
This model differs from the ELMM because it uses only one factor per pixel and not one per component per pixel. Consequently, the simplex can shift but stays parallel to the initial one with our model, whereas in [7], the simplex can shift anyhow as long as it stays in the cone generated in $\mathbf{A}$. We point out that since our data have a high spatial resolution, the pixels are mostly pure. Therefore, the samples belong to the edges of the cone, for which both models are equivalent. The main benefits of the simplified model are that its computational load is more manageable and it is also not ill-posed (each sample is associated with a unique factorization with our model but an infinity with ELMM without the spatial regularization). In contrast, the ELMM handles more correctly boundaries between background materials, especially if the energies of the background spectra in $\mathbf{A}$ do not correspond exactly to the observed ones.
In many applications, the intrinsic energy of the components and global illumination of the area are more constant than the abundances. Therefore, the assumption on the spatial regularity of the energy factors is more likely to be valid than the one on the abundances, especially with if the spatial resolution is not really high. This modelling should be then more efficient than the previous cone model with constraints on the abundances.

In order to increase the robustness of this approach, we keep the weighting scheme proposed previously for the cone model. We estimate the background by solving the following problem:

$$\underset{\mathbf{S}\in\mathcal{S}^{n-1},\mathbf{\Lambda}}{\operatorname{argmin}} \frac{1}{2} \left\|(\mathbf{X}-\mathbf{AS\Lambda})\mathbf{W}_O\right\|_2^2 + \alpha \left\|\mathbf{\Lambda}\Phi\right\|_1 \qquad (2)$$

Matrix Φ denotes the transformed domain in which the energy factors are sparse.

The algorithm is described as the following: i) estimate $\tilde{\mathbf{S}}$ by assuming that it is non-negative, ii) initialize $\Lambda_{k,k} = \sum_{i=1}^{n} \tilde{\mathbf{S}}_i^k$, $\forall k = 1..t$ and $\mathbf{W}_O$, iii) estimate jointly $\mathbf{S}$ and Λ alternatively with (2) [15], with a Forward-Backward (FB) for $\mathbf{S}$ [16] and a Condat-Wu [14] for Λ (or FB if Φ is orthonormal).

4. TEST ON REAL DATASETS

We establish our comparisons on two datasets provided by the UK Defence Science and Technology Laboratory (Dstl); see Fig.3. The data come from the Selene trial, which collected airborne hyperspectral imagery of large numbers (hundreds) of spectrally varied targets across a two week period [1]. One scene is composed of 400×200 pixels in radiance in the VNIR range with 140 bands, with a resolution of 25×25cm. It is assumed that the background is composed of 5 components, two kinds of soil, road, grass and tree, whose spectra have been manually extracted and also used in [1]. Thirty man-made targets, made of either green or grey ceramic, exist in the scene. The second scene is composed of 310×170 pixels at another location and in the reflectance mode. Background spectra (road, soil and two types of grass) have been estimated with VCA [17] at another sunny location (the choice of $\mathbf{A}$ has a significant influence on the results). It is composed of 25 targets of beige carpet which are particularly challenging to retrieve: the spectrum of the beige carpet is highly correlated with the one of the soil and in contrast with the ceramics, its energy is similar to the background. The illumination is not constant.

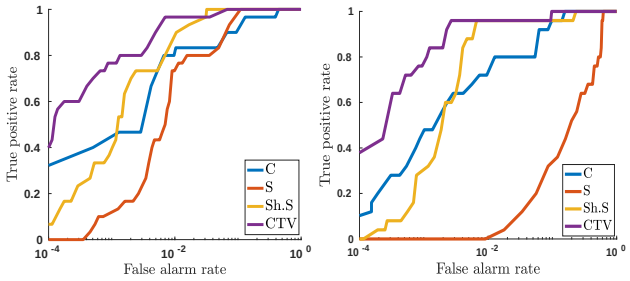
We will estimate the background by solving the following problems:

- (S): Simplex model: $\operatorname{argmin}_{\mathbf{S}\in\mathcal{S}^{n-1}} \left\|\mathbf{X}-\mathbf{AS}\right\|_2^2$.
- (C): Cone model $\operatorname{argmin}_{\mathbf{S}\geq 0} \left\|\mathbf{X}-\mathbf{AS}\right\|_2^2$.
- (CTV): Cone model and TV norm on the sources in eq.(1)
- (Sh.S): Proposed shifting simplex in eq.(2) with the TV norm.

[1] Data from the Selene trial are available through the UDRC. Please contact the UDRC data manager at UDRC-Datacentre@dstl.gov.uk.

(a) Data 1 (b) Data 2

Fig. 3: False color images of the two datasets.



(a) ROC Data 1 (b) ROC Data 2

Fig. 4: False colour images of the two datasets.

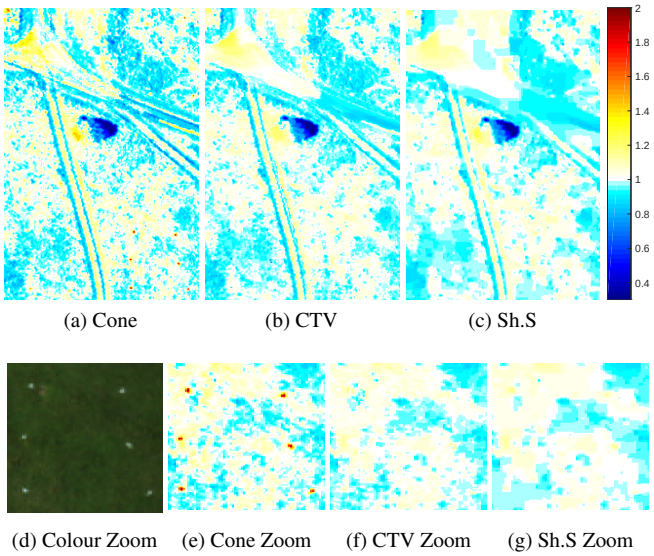
The support of $\mathbf{O}$ is obtained from the residual $\mathbf{X} - \mathbf{AS}$: an outlier is present if the energy of the residual, at a given pixel, is greater than β. This parameter β will vary and we will compute the number of detected targets (True positive) and the number of false positive detections (False alarm) for each value. The ROC curves will be displayed for comparison between the methods.

The joint estimation of the outliers and the abundances would not affect the support of the outliers and so the results of (S) and (C). The results should be improved for (Sh.S) and (CTV) but not significantly due to the weighting scheme.

4.1. Results

The ROC curves for the two datasets are displayed in Fig.4. In the first data set, the methods (CTV) and (Sh.S), which exploit the spatial information, are efficient to detect the non-obvious targets. The cone model is also quite robust but it struggles to estimate some of the targets, *i.e.* high FA for achieving 1 true positive rate. On the other hand, the simplex method leads to a higher FA on average due to the shadow of the tree, but more easily detects the smallest targets by using the energy variations. The energy factors are presented in Fig.5: we can see that (CTV) and (Sh.S) can indeed recover an energy map which is not affected by the targets, while still taking into account the energy changes such as the shadow of the tree or slightest intrinsic energy variations between 0.8 and 1.2 as shown in the maps Fig.5.

Similar remarks hold for the second dataset. Once again, (CTV) and then (Sh.S) are more efficient. The simplex method (S) performs poorly because the illumination of the scene is not constant. For both approaches (simplex and cone), exploiting spatial information clearly improves the results. The ones of (Sh.S) could be improved by using a regularization better allowing for smooth variations (the TV norm prefers sharp variations).



(a) Cone (b) CTV (c) Sh.S

(d) Colour Zoom (e) Cone Zoom (f) CTV Zoom (g) Sh.S Zoom

Fig. 5: Estimated energy factors: $\sum_{i=1}^{n} \mathbf{S}_i$ for (C), (CTV) and Λ for (Sh.S). Top row: full area, first dataset. Bottom row: zoom on the area with grey ceramic targets (false colour image and then energy factor). The background energy factors have been fairly retrieved in (f) and (g).

As a summary, (Sh.S) and (CTV) are more efficient than the pixel-wise methods. In these two examples, (CTV) was more precise because the spatial resolution of the data was really high. (Sh.S) may be more efficient with a lower spatial resolution.

5. CONCLUSION

We propose two background models which benefit from the high spatial resolution of the data to spatially constrain the background. We can then exploit jointly the energy and spectral variations created by the targets while being robust to energy variations of the background. The preliminary tests on real data support this approach. In practice, spectral variabilities can be more complex than the scaling effects considered here [18]. Modelling these variabilities is necessary to maintain the false alarm rate. It is particularly challenging within the anomaly detection framework since we need to design a model able to synthesis these background spectral variabilities but not the variations created by the targets. This is left to future work.

6. REFERENCES

[1] Abderrahim Halimi, Yoann Altmann, Gerald S Buller, Steve McLaughlin, William Oxford, Damien Clarke, and Jonathan Piper, "Robust unmixing algorithms for hyperspectral imagery," in *Sensor Signal Processing for Defence (SSPD), 2016*. IEEE, 2016, pp. 1–5.

[2] Telmo Adão, Jonáš Hruška, Luís Pádua, José Bessa, Emanuel Peres, Raul Morais, and Joaquim João Sousa, "Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry," *Remote Sensing*, vol. 9, no. 11, pp. 1110, 2017.

[3] Nirmal Keshava and John F Mustard, "Spectral unmixing," *IEEE signal processing magazine*, vol. 19, no. 1, pp. 44–57, 2002.

[4] Abderrahim Halimi, José M Bioucas-Dias, Nicolas Dobigeon, Gerald S Buller, and Stephen McLaughlin, "Fast hyperspectral unmixing in presence of nonlinearity or mismodeling effects," *IEEE Transactions on Computational Imaging*, vol. 3, no. 2, pp. 146–159, 2017.

[5] Y. Altmann, S. McLaughlin, and A. Hero, "Robust Linear Spectral Unmixing Using Anomaly Detection," *IEEE Transactions on Computational Imaging*, vol. 1, no. 2, pp. 74–85, June 2015.

[6] C. Fevotte and N. Dobigeon, "Nonlinear Hyperspectral Unmixing With Robust Nonnegative Matrix Factorization," *Image Processing, IEEE Transactions on*, vol. 24, no. 12, pp. 4810–4819, Dec 2015.

[7] Lucas Drumetz, Miguel-Angel Veganzones, Simon Henrot, Ronald Phlypo, Jocelyn Chanussot, and Christian Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3890–3905, 2016.

[8] Abderrahim Halimi, Nicolas Dobigeon, and Jean-Yves Tourneret, "Unsupervised unmixing of hyperspectral images accounting for endmember variability," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4904–4917, 2015.

[9] C. Chenot and J. Bobin, "Blind source separation with outliers in transformed domains," *SIAM Journal on Imaging Sciences*, vol. 11, no. 2, pp. 1524–1559, 2018.

[10] Marian-Daniel Iordache, José M Bioucas-Dias, and Antonio Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, 2012.

[11] Qiangqiang Yuan, Liangpei Zhang, and Huanfeng Shen, "Hyperspectral image denoising employing a spectral–spatial adaptive total variation model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3660–3677, 2012.

[12] Jean-Luc Starck, Fionn Murtagh, and Jalal M Fadili, *Sparse image and signal processing: wavelets, curvelets, morphological diversity*, Cambridge University Press, 2010.

[13] Huanfeng Shen, Xinghua Li, Qing Cheng, Chao Zeng, Gang Yang, Huifang Li, and Liangpei Zhang, "Missing information reconstruction of remote sensing data: A technical review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 61–85, 2015.

[14] Laurent Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, 2013.

[15] Paul Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of optimization theory and applications*, vol. 109, no. 3, pp. 475–494, 2001.

[16] Neal Parikh, Stephen P Boyd, et al., "Proximal algorithms," *Foundations and Trends in optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[17] José MP Nascimento and José MB Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.

[18] Alina Zare and KC Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 95–104, 2014.